JOURNAL OF BACTERIOLOGY, Jan. 1994, p. 265–267
0021-9193/94/$04.00+0

Vol. 176, No. 1

Uninterrupted Translation through Putative 12-Nucleotide Coding Gap in Sequence of *carA*: Business as Usual

THÉRÈSE M. F. TUOHY,[1]* THOMAS KIDD,[1]† RAYMOND F. GESTELAND,[1,2] AND JOHN F. ATKINS[1]

Department of Human Genetics[1] and Howard Hughes Medical Institute,[2] 6160 Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah 84112-1107

Received 9 September 1993/Accepted 30 October 1993

Previous work of others reported an untranslated stretch of 12 nucleotides in the 5′ coding sequence of *carA* from *Pseudomonas aeruginosa*. However, N-terminal protein sequencing of *carA-lacZ* translational fusions shows that these 12 nucleotides are normally translated in a continuous triplet manner, both in *P. aeruginosa* and in *Escherichia coli*.

In recent years, exceptions have been discovered to the previously universal rule of one amino acid for one codon, in strictly sequential, uninterrupted, nonoverlapping order. It is now relatively common to find a gene whose expression involves translational read-through, frameshifting, or other translational anomalies. The more intriguing of these anomalies may autogenously control the ratio of normal to frameshifting translation (1). Several of these genes share an apparently common and elegant duality in design of these sequences; each of them consists of a frameshifting site, and a nearby activator which appears to operate by transiently modulating the translational fidelity of the ribosome across the slippery site (5).

The majority of known examples of deviations from normal translation involve stop codon read-through or reading frameshifts of +1 or −1. A few low-level hops have also been reported, and these underline a previously undetected possible source of ribosome flexibility in normal translation (11, 18). By contrast, the most unorthodox example of aberrant translation involves a jump of +50 nucleotides in gene *60* of phage T4 (6). The factors that contribute to this shift have been determined to be (i) an interval of approximately 50 nucleotides, bordered at the 5′ and 3′ junctions by identical codons, whose identity per se does not appear to be critical; (ii) a stop codon at the immediate 5′ end of the coding gap, which forms part of (iii) a strong stem loop; and (iv) a lysine-arginine-rich nascent peptide translated from sequences just 5′ to the junction (19). A second example of hopping, over a rare arginine codon, has been reported for the expression of a bovine gene in *Escherichia coli* (7). There is a third reported case, also in *E. coli*, of a 55-nucleotide hop in a *trpR-lacZ* translational fusion (2), but the contributing elements remain to be deciphered.

A possibly related, in-frame hop has been reported for *carA* (encoding the α subunit of carbamoylphosphate synthetase) of *Pseudomonas aeruginosa* for which protein sequencing clearly showed the absence of four expected amino acids (20). As the cDNA sequence reflected the genomic DNA sequence, including 12 nucleotides not represented in the protein sequence (20), it seemed likely that the translation of *carA* involves the translational occlusion of 12 nucleotides (Fig. 1). mRNA splicing (for which there is no comparative sequence evidence) appears to be extremely unlikely without added levels of complexity. However, the possibility of protein splicing, while not considered, cannot be excluded, as there are now some precedents (3), albeit not in prokaryotes.

Because of the existence of short direct and inverted repeats in this region, it is not possible to determine precisely which 12 nucleotides between codons 4 and 9, present in the mRNA sequence, are not represented in the resulting protein (Fig. 1a). Consideration of these extensive direct and inverted repeats suggests that it is possible to hypothesize either of two secondary structures which are mutually exclusive, and therefore compete with each other. They can be thought of as isomers of each other, and illustrated as a triple helix (Fig. 1b). (There are reports of extended tertiary interactions [14] in mRNA [10] and tRNA [13].) CCA proline codons 4 and 8 fall in the two direct repeats, and their base-pairing complement, UGG, falls into an inverted repeat, which has the flip-flop option of base-pairing with either. This structure might provide an activation template for the translating ribosome to pirou-

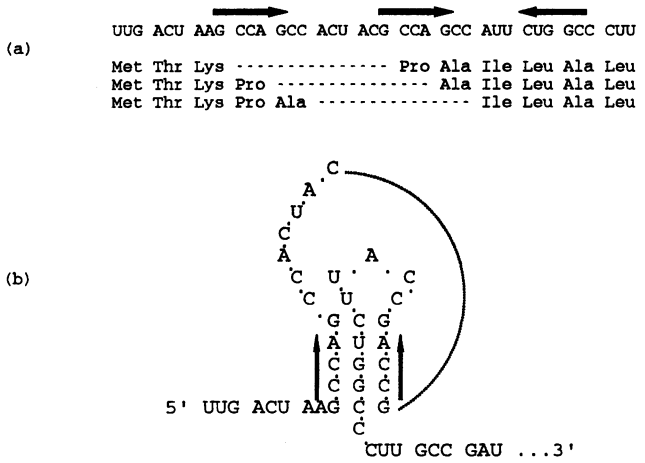

FIG. 1. (a) Three alternative possibilities for 12-nucleotide hop in translation of *carA* mRNA, given the direct repeats, and the results of N-terminal protein sequencing (20). (b) Alternative stem-loops of the 5′-coding sequence of *carA* drawn in the form of a triple helix (see text). Direct and inverted repeats are indicated by arrows.

* Corresponding author. Mailing address: Department of Human Genetics, 6160 Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84112-1107. Phone: (801) 581-4429. Fax: (801) 585-3910. Electronic mail address: therese@genetics.med.utah.edu.

† Present address: Developmental Genetics Laboratory, I.C.R.F. Developmental Biology Unit, Department of Zoology, Oxford OX1 3PS, England.

```
P.a.   TTG ACT AAG CCA GCC ACT ACG CCA GCC ATT CTG GCC CTT GCC
E.c.   --- -T- --- T-- --G                 C-A T-- -TT --G -AA
S.t.   --- -T- --- T-- --G                 C-A T-- -TT --G -AA

P.a.   GAT GGC AGC ATC TTT CGC GGC GAA TCC ATC GGC GCC GAT GGC
E.c.   --C --A -C- CAG --- -A- --T CGG G-- --A --G --A ACA --T
S.t.   --C --A -C- CAT --- -A- --T CGG G-- --T --G --A ACG --T
```

FIG. 2. Alignment of the 5′ 84 nucleotides of the *carA* coding region of *P. aeruginosa* (*P.a.* [20]) with the same region in *E. coli* (*E.c.* [12]) and *S. typhimurium* (*S.t.* [8]). The underlined nucleotides are those not represented in the reported protein sequence (20).

ette from the first to the second proline codon, thereby spanning exactly 12 nucleotides. Such unmelted secondary structure surviving the processivity of the translating ribosome was unprecedented, but examples of secondary structure within the span of the ribosome critically influencing translation are emerging (19).

It is particularly intriguing that the best alignment of *carA* open reading frames from *E. coli*, *Salmonella typhimurium*, and *P. aeruginosa* is obtained by disregarding 12 nucleotides in the region of the reported gap (Fig. 2). This alignment produces identity at 30 of 57 bases immediately following the 12-nucleotide stretch (compared with an expected 14 to 15 on the basis of random chance), while a straight alignment (without looping out 12 bases of *P. aeruginosa*) gives identity at only 10 of 57.

Initial experiments were designed to reproduce the original report, before attempting to identify the contributing component(s). With primers based on sequences 25 or 250 nucleotides upstream and 124 nucleotides downstream of the *carA* initiation codon, segments of *carA* were amplified from pSW4 (20) by PCR and cloned into the 5′ terminus of the *lacZ* gene of a derivative of pBR322, between the genetically engineered *Xba*I and *Apa*I sites (15, 18). Resulting constructs were introduced into *E. coli* XL1-Blue (Stratagene). N-terminal protein sequencing of β-galactosidase fusion protein expressed and purified (9) from two of these constructs showed normal continuous translation through the reported gap, and the protein sequence obtained from the longer of these is shown in Fig. 3a. This was a surprise, as previous attempts to reproduce translational irregularities by the creation of convenient reporter gene fusions have been successful for many systems, perhaps most notably gene *60* of phage T4 (6, 19), with technology identical to that described here.

To more closely approximate the wild-type conditions of the reported hop, it was decided to (i) clone the entire *carA* upstream untranslated transcript and (ii) use *P. aeruginosa* PAO1 as the host strain. (There is no known precedent for a prokaryotic gene whose internal translational control depends on more distant sequences.) Therefore, primers based on the sequences 500 nucleotides upstream and 240 nucleotides downstream of the *carA* TTG initiation codon were used to amplify *carA*, resulting in a 746-nucleotide fragment which was cloned into the *Hin*dIII and *Eco*RI sites of the multiple cloning site of the gram-negative shuttle vector pQF50 (4), creating a *carA-lacZ* translational fusion. Resulting constructs were initially cloned into *E. coli* and transferred by electroporation into *P. aeruginosa* PAO1. Protein sequencing showed that translation of the *carA* mRNA is continuous in *P. aeruginosa* and did not reveal an untranslated region of 12 nucleotides at the reported position (Fig. 3b).

As the limit of resolution of a theoretical secondary sequence is estimated to be 10% of the primary sequence, it was possible that the reported hop occurs in the *carA-lacZ* fusions,

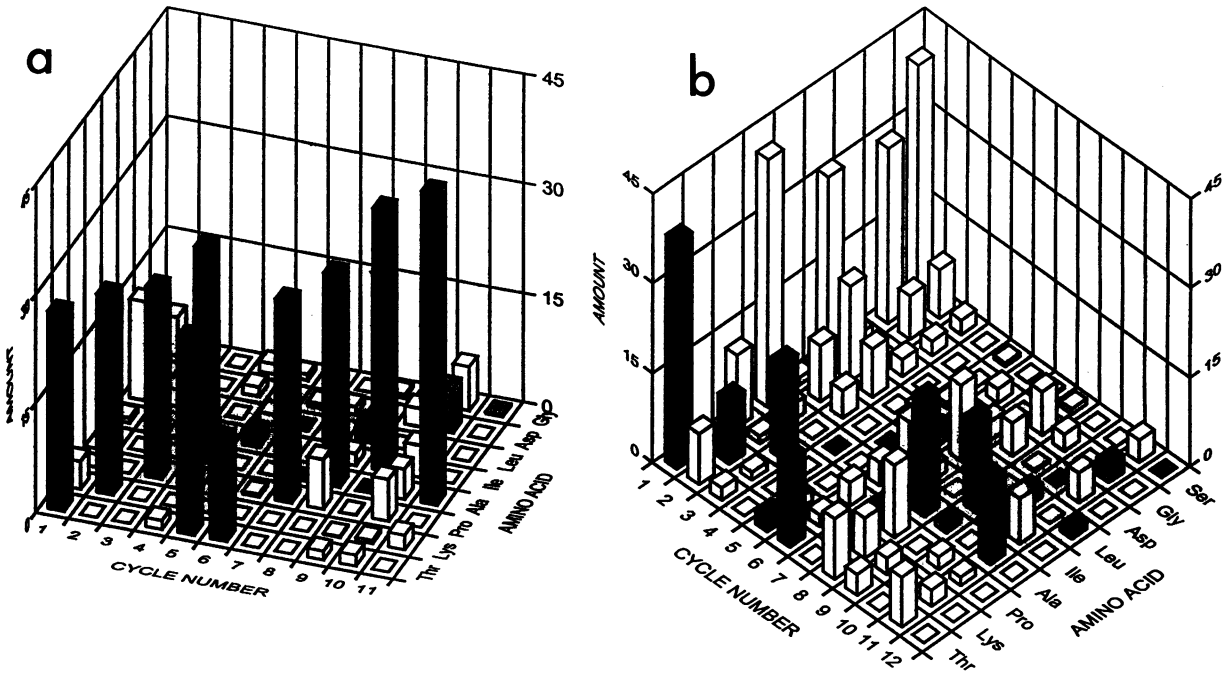

FIG. 3. (a) Graphic representation of N-terminal sequencing of *carA-lacZ* translational fusion which included ~250 nucleotides upstream and ~120 nucleotides downstream of the *carA* initiation codon cloned into a derivative of pBR322 and expressed in *E. coli*. (b) N-terminal sequencing of *carA-lacZ* translational fusion which included ~500 nucleotides upstream and ~240 nucleotides downstream of the *carA* initiation codon, cloned into pQF50, and expressed in *P. aeruginosa*. The heavily shaded bars show the sequence expected from uninterrupted triplet translation, the lightly shaded bars show the sequence expected in the absence of the relevant four amino acids, and the white bars show the background amino acids. Computer graphic representation was obtained by use of the Deltagraph Professional software package (DeltaPoint).

but at a low efficiency. To address this possibility, a construct was designed in which the sixth ACU threonine codon was replaced by a UAA stop codon, by PCR mutagenesis (16), so that any low-level hopping event might be revealed in the absence of normally translated mRNA. On the basis of the sequence of the region, this is the most conservative codon to mutate, as it is the only one which can be deduced to necessarily be included in the unrepresented nucleotides, regardless of which 12-nucleotide span constitutes the gap (Fig. 1a). Expression of the resulting clone failed to produce any measurable β-galactosidase, as judged by enzyme assays, making protein sequence inapplicable. However, as the mechanism of action of the hop in the translation of the native mRNA remains to be understood, the ACU codon could very well be a requirement, and so no positive deductions can be made from the failure of this clone to reveal the hop.

It is possible that a required, *trans*-acting factor is diluted beyond critical concentration for the detection of the bypass, because of the relatively high copy number of the vectors (~30 in *E. coli* for the pBR322-derived constructs and ~13 in *P. aeruginosa* for the pQF50-derived construct). However, there is no a priori evidence for such a factor, and its existence is purely speculative.

It is also possible that a *cis*-acting element was not included in the cloning of the gene fusions. This seems unlikely, however, as the entire 500 nucleotides of upstream untranslated sequence and 240 nucleotides of downstream sequence (with respect to the TTG initiation codon) were included in the largest clone. However, there may be something special about the genomic context of *carA*, such that nearby sequences play a role in the hopping event, perhaps through the coupling of transcription and translation.

In spite of the apparent cotranslational hopping over four consecutive codons of the *carA* mRNA in *P. aeruginosa* (20), several attempts to reproduce this phenomenon in *carA-lacZ* fusions have failed. *E. coli* strains were grown in Luria-Bertani broth (tryptone, 10 g/liter; yeast extract, 5 g/liter; NaCl, 5 g/liter), and *P. aeruginosa* was grown in citrate minimal medium (17) at 37°C as reported for the native hop. We can only conclude that unless very special, and subtle, conditions are required, that translation of *carA* of *P. aeruginosa* through the 12 nucleotides of interest proceeds according to the established rules of triplet translation.

We thank A. Abdelal for pSW4, M. Farinha for shuttle vector pQF50, and J. Jesse for *P. aeruginosa* PAO1. We also thank Ed Meenen for expert help with protein sequencing, and Bob Weiss for critical reading of the manuscript.

This work was supported by a research grant from NIH (R01-GM48152) and by the Howard Hughes Medical Institute.

REFERENCES

1. **Atkins, J. F., R. B. Weiss, and R. F. Gesteland.** 1990. Ribosome gymnastics—degree of difficulty 9.5, style 10.0. Cell **62:**413–423.
2. **Benhar, I., and H. Engelberg-Kulka.** 1993. Frameshifting in the expression of the *E. coli trpR* gene occurs by the bypassing of a segment of its coding sequence. Cell **72:**121–130.
3. **Copper, A. A., and T. H. Stevens.** 1993. Protein splicing: excision of intervening sequences at the protein level. Bioessays **15:**667–674.
4. **Farinha, M. A., and A. M. Kropinski.** 1990. Construction of broad-host-range plasmid vectors for easy visible selection and analysis of promoters. J. Bacteriol. **172:**3496–3499.
5. **Gesteland, R. F., R. B. Weiss, and J. F. Atkins.** 1992. Recoding: reprogrammed genetic coding. Science **257:**1640–1641.
6. **Huang, W. M., S.-W. Ao, S. Casjens, R. Orlandi, R. Zeikus, R. B. Weiss, D. Winge, and M. Fang.** 1988. A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene *60*. Science **239:**1005–1012.
7. **Kane, J. F., B. N. Violand, D. F. Curran, N. R. Staten, K. L. Duffin, and G. Bogosian.** 1992. Two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of *Escherichia coli*. Nucleic Acids Res. **20:**6707–6712.
8. **Kilstrup, M., C. Lu, A. Abdelal, and J. Neuhard.** 1988. Nucleotide sequence of the *carA* gene and regulation of the *carAB* operon in *Salmonella typhimurium*. Eur. J. Biochem. **176:**421–429.
9. **Matsudaira, P.** 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. **262:**10035–10038.
10. **Michel, F., A. D. Ellinton, S. Couture, and J. W. Szostak.** 1990. Phylogenetic and genetic evidence for triple bases in catalytic domain of group I introns. Nature (London) **347:**578–580.
11. **O'Connor, M., R. F. Gesteland, and J. F. Atkins.** 1989. tRNA hopping: enhancement by an extended anticodon. EMBO J. **8:**4315–4323.
12. **Piette, J., H. Nyunoya, C. J. Lusty, R. Cunin, G. Weyens, M. Crabeel, D. Charlier, N. Glansdorff, and A. Piérard.** 1984. DNA sequence of the *carA* gene and the control region of *carAB*: tandem promoters, respectively controlled by arginine and the pyrimidines, regulate the synthesis of carbamoyl-phosphate synthetase in *Escherichia coli* K12. Proc. Natl. Acad. Sci. USA **81:**4134–4138.
13. **Puglisi, J. D., J. R. Wyatt, and I. Tinoco, Jr.** 1990. Conformation of a pseudoknot. J. Mol. Biol. **214:**437–453.
14. **Roberts, R. W., and D. M. Crothers.** 1992. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science **258:**1463–1466.
15. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Shyamala, V., and G. F.-L. Ames.** 1991. Use of exonuclease for rapid polymerase-chain-reaction-based in vitro mutagenesis. Gene **97:**1–6.
17. **Stalon, V., F. Ramos, A. Piérard, and J. M. Wiame.** 1967. The occurrence of a catabolic and an anabolic ornithine carbamoyltransferase in Pseudomonas. Biochim. Biophys. Acta **139:**91–97.
18. **Weiss, R. B., D. M. Dunn, J. F. Atkins, and R. F. Gesteland.** 1987. Slippery runs, shifty stops, backward steps, and forward hops: −2, −1, +1, +2, +5 and +6 ribosomal frameshifting. Cold Spring Harbor Symp. Quant. Biol. **52:**687–693.
19. **Weiss, R. B., D. M. Dunn, J. F. Atkins, and R. F. Gesteland.** 1990. Ribosomal frameshifting from −2 to +50 nucleotides. Prog. Nucleic Acid Res. Mol. Biol. **39:**159–183.
20. **Wong, S. C., and A. T. Abdelal.** 1990. Unorthodox expression of an enzyme: evidence for an untranslated region within *carA* from *Pseudomonas aeruginosa*. J. Bacteriol. **172:**630–642.